

High-level visual object representations in inferior temporal cortex

Citation for published version (APA):

Mur, M. C. (2011). *High-level visual object representations in inferior temporal cortex*. [Doctoral Thesis, Maastricht University]. Maastricht University Press. <https://doi.org/10.26481/dis.20111202mm>

Document status and date:

Published: 01/01/2011

DOI:

[10.26481/dis.20111202mm](https://doi.org/10.26481/dis.20111202mm)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

We live in a rich visual environment, which is populated by many different kinds of objects. Recognition of these objects is essential for successful interaction with our environment. Object recognition is a computationally challenging task, and is accomplished by the primate brain in several processing steps along the ventral visual pathway. These processing steps result in high-level object representations at the level of inferior temporal (IT) cortex, which are fairly invariant to image transformations and form the basis for categorization and higher-order cognitive processes. These high-level object representations in IT are the focus of this thesis.

IT object representations have been studied in both humans and monkeys. While neurophysiological studies in monkeys investigated IT responses to individual object exemplars, neuroimaging studies in humans only assessed category-average IT responses. As a consequence, little is known about human brain responses to individual object exemplars. Furthermore, quantitative comparison of human and monkey data has been complicated by the need for defining the correspondence between measurement units. Similar correspondence problems complicate relating brain data to computational theory and behavior, hampering the development of an integrated theory of vision, and of a unified systems neuroscience in general. The work presented in this thesis addressed these issues by (1) assessing human IT responses to real-world object exemplars using fMRI, and (2) comparing the measured responses to data from monkey IT, computational models, and human behavior using the newly developed framework of representational similarity analysis (RSA). A summary of each chapter is given below, followed by a concise overall summary and discussion of the main findings.

Chapter summaries

The different fMRI analysis methods used in the presented work are described and compared in **Chapter 1**. This chapter focuses on pattern-information analysis, which in recent years has gained momentum in the field of neuroimaging. Pattern-information analysis aims to detect multivoxel activity-pattern differences between experimental conditions. These differences can be interpreted as reflecting differences in underlying neuronal population activity which is thought to represent mental content. Another technique that targets representational content is fMRI adaptation. This technique is based on the logic that fMRI stimulus-change effects in a specific brain region can be interpreted to indicate that the region contains neurons that represent the changed stimulus property.

fMRI adaptation has been widely used in the field of face perception to infer the presence of face-identity representations. Previous studies have suggested the

fusiform face area (FFA) as the locus of face-identity representations, but did not thoroughly investigate (1) the specificity of the effects to FFA and (2) the effect of low-level stimulus changes on the spatial extent of effects. **Chapter 2** reports that face-identity-change effects are not confined to face-selective regions: effects were also found in early visual cortex and the parahippocampal place area (PPA). To ensure that face-identity-change effects could not be attributed to low-level stimulus changes, we introduced viewpoint and illumination changes on both face-identity repetition and face-identity change trials. This led to a decrease in the spatial extent of effects, but did not eliminate effects outside of face-selective regions. Our findings could be interpreted as evidence for high-level face-identity representations in early visual cortex and PPA, but this seems unlikely given the known response properties of these regions. Alternatively, our effects could be explained by general attentional effects or carryover of activation from connected regions. Our results suggest that fMRI stimulus-change effects do not provide conclusive evidence for a neuronal representation of the changed stimulus property.

We therefore abandoned the fMRI-adaptation approach and moved to the more direct approach of simply measuring fMRI responses to individual object images using ungrouped-events designs (i.e. each object image is treated as a separate condition). Ungrouped-events designs feature in representational similarity analysis (RSA), which is introduced in **Chapter 3**. RSA is a new experimental and data-analytical framework that enables quantitative comparison of data from different branches of systems neuroscience by abstracting from activity patterns and relating the data at the level of similarities between activity patterns. Comparison of represented information at the level of activity-pattern similarity obviates the need for defining the correspondency between measurement units. Potential applications of RSA include validation of computational models by brain-activity data, relating brain representations between different species, analyzing representational connectivity between different brain regions, and exploring the representational content of brain regions using condition-rich ungrouped-events designs. Statistical inference is performed using randomization and bootstrap techniques. RSA was demonstrated by relating representations of real-world object images measured by fMRI in early visual cortex and FFA to a range of computational models. Consistent with existing literature, the object representation in early visual cortex was best explained by a simple silhouette-image model, and the representation in FFA was best explained by a conceptual face-animal-prototype model. Before moving to implementations of RSA in Chapters 5 and 6, we turn to Chapter 4.

Chapter 4 bridges a gap between classical fMRI analysis and RSA. Classical fMRI studies on object representation in IT investigated category-average activation of brain regions. RSA investigates object-exemplar pattern-information, incor-

porating two advances at once: from category-average to exemplar responses and from activation to pattern-information analysis. This creates two “gaps”. The first gap, category-average pattern-information analysis, has been addressed previously. What is missing is the analysis of activation of category-selective brain regions to individual object exemplars. In other words, would category-selectivity hold for individual objects? Bridging this second gap also establishes a link to neurophysiological studies in monkeys, which commonly measure neuronal activity in response to individual object exemplars. We measured fMRI activation of category-selective regions FFA and PPA to 96 real-world object images from a wide range of categories, including faces and places. We found no evidence of any images outside the preferred category eliciting a stronger response than any images inside the preferred category for either FFA or PPA. Regional-average activation might thus perfectly reflect category membership of individual objects. Within each category, individual images elicited different levels of activation, suggesting a graded rather than a pure step-function response profile. The 96-object fMRI data were subsequently analyzed for pattern-information within the RSA framework in the last two chapters.

Chapter 5 relates the IT representations of the same object images between human (fMRI) and monkey (cell recording). We found that IT activity patterns cluster according to natural categories: animate and inanimate objects formed top-level category clusters; faces and bodies formed subclusters within the animates. This hierarchical categorical structure inherent to IT matched between man and monkey. Within-category exemplar similarities also matched between the species. Results were robust against exclusion of category-selective regions FFA and PPA from analysis. Species-specific face analysis suggested an exception to the close match between man and monkey: IT better distinguished conspecific faces in each species. A range of low- and intermediate-level computational models could not account for the categorical representation observed in IT, indicating that our results cannot be explained by low-level feature similarity alone. In sum, these findings suggest that primate IT across species may host a common code, which combines a categorical and a continuous representation of objects.

The presence of IT activity-pattern clusters that correspond to well-known object categories suggests a link to human perception. This link was investigated in **Chapter 6** by relating the IT object representation to human object-similarity judgments of the same 96 real-world object images. Given our relatively large stimulus set, conventional methods for obtaining pairwise similarity judgments would be very time-consuming. We therefore developed a new multi-arrangement method for efficient and subject-tailored measurement of perceived similarity for large sets of stimuli. We found that human similarity judgments cluster the objects by category and reflect several features of the primate

IT object representation, including the basic distinction between animate and inanimate objects. Within-category exemplar similarities also matched between similarity judgments and IT object representations. These findings suggest that high-level conscious similarity judgments of real-world object images reflect visual similarities and categorical distinctions of longstanding evolutionary relevance that are represented at the level of primate IT. However, human similarity judgments transcended the IT representational stage in terms of a stronger categorical component and the introduction of species-dependent (human/nonhuman) and evolutionarily recent (natural/artificial) distinctions. These additional distinctions may reflect a prefrontal contribution allowing more flexible categorical distinctions.

Overall summary and discussion

Our findings suggest that category membership of individual objects is an important organizing principle of the primate IT object representation. This idea is not new: previous studies have shown that IT responses contain information about category membership (e.g. Tanaka, 1996; Tsao et al., 2006; Puce et al., 1995; Kanwisher et al., 1997; Haxby et al., 2001). However, we go beyond these studies using ungrouped-events designs and RSA, showing that the structure of the IT object representation is inherently categorical and hierarchically organized, with clusters reflecting object categories of longstanding evolutionary relevance that match between man and monkey. These “primate-default” categories represented at the level of IT were also reflected in high-level conscious human object-similarity judgments. A range of computational models containing low- and intermediate-level object representations could not explain the categorical structure observed in IT. Our findings suggest that IT models could be improved by the acquisition of category-discriminating features through supervised learning (Ullman et al., 2002; Sigala and Logothetis, 2002).

Our results indicate that IT does not only distinguish categories, but also individual object exemplars within each category. This finding is consistent with previous findings in the monkey (e.g. Young and Yamane, 1992; Hung et al., 2005) and with recent reports of pattern-information differences between exemplars of the same category in human IT (Kriegeskorte et al., 2007; Eger et al., 2008). We found that the within-category similarity representations match between species, and between brain data and similarity judgments. The within-category match is likely driven by visual similarities and dissimilarities between objects that belong to the same category cluster, and is consistent with the previously reported relationship between perceived object shape and primate IT activity-pattern similarities (Edelman et al., 1998; Haushofer et al., 2008; Op de Beeck et al., 2001; 2008).

The tightest activity-pattern cluster in both IT and FFA was formed by human faces, likely reflecting the high visual similarity between different face identities. Consistent with this observation, a previous attempt to detect face-identity pattern information in FFA failed; face-identity information was detected in anterior IT instead (Kriegeskorte et al., 2007). Cell recording from the middle macaque face patch, a suggested homologue of FFA, indicated some amount of face-identity information in its population response (Tsao et al., 2006), but the category effects explained most of the response variance. A recent fMRI study in humans detected face-identity information in anterior FFA and IT using spatio-temporal pattern analysis (Nestor et al., 2011). Our face-identity change findings would be consistent with this finding, but cannot provide conclusive evidence for the existence of face-identity representations in FFA. In sum, the current evidence suggests that face recognition is performed by a network of regions, including FFA and anterior IT. FFA might carry some amount of face-identity information, but its main function seems to be face detection (Puce et al., 1995; Kanwisher et al., 1997; Kriegeskorte et al., 2007).

In sum, our findings suggest that primate IT may host a common code across species, which combines a categorical and a continuous representation of objects. This code might be implemented by a continuous feature map containing several high-density clusters of related features tuned for the discrimination of categories of high behavioral and evolutionary relevance (e.g. faces) (Haxby et al., 2001). Such a map could explain the existence of category-selective regions, and would be consistent with our finding that activation of these regions appears to perfectly reflect category membership of individual objects, but nevertheless follows a graded instead of a pure step-function response profile. Furthermore, such a map would be consistent with our finding that activity patterns form clusters corresponding to conventional object categories despite the absence of step-function-like categorical responses in IT at the single-cell or single-voxel level. Step-function-like categorical single-cell responses have been reported for prefrontal cortex (Ashby and Ell, 2001; Freedman et al., 2001). Prefrontal cortex receives input from IT and might combine the information distributed in IT to explicate categories in a flexible task-dependent manner. This region might, in combination with IT, also contribute to high-level conscious human object-similarity judgments. Avenues for future research include investigation of the nature of the feature map(s) in IT using high-field fMRI (see Op de Beeck et al., 2008) and exploration of the representational connectivity between IT and prefrontal cortex within and across species using RSA.

Samenvatting

We leven in een rijke visuele omgeving waarin zich veel verschillende voorwerpen bevinden. Herkenning van deze voorwerpen is essentieel om succesvol te kunnen interacteren met onze omgeving. Het herkennen van voorwerpen is computationeel gezien een uitdagende taak. Deze taak wordt door het brein uitgevoerd in verschillende verwerkingsstappen in het ventrale visuele systeem. De verwerkingsstappen resulteren in hogere-orde representaties van voorwerpen op het niveau van de inferieure temporele (IT) cortex. Deze representaties zijn redelijk bestand tegen transformaties van de binnenkomende visuele informatie en vormen de basis voor het categoriseren van voorwerpen en andere hogere-orde cognitieve processen. Genoemde hogere-orde representaties van voorwerpen in IT vormen het onderwerp van dit proefschrift.

Representaties van voorwerpen in IT zijn onderzocht in zowel mensen als apen. In neurofysiologisch onderzoek in apen is het gebruikelijk om hersenresponsen op individuele voorwerpen te meten. In onderzoek met beeldvormende technieken in mensen is dit niet gebruikelijk: hersenresponsen worden over het algemeen gemeten voor een set van voorwerpen, waardoor de data de hersenactiviteit gemiddeld over verschillende voorwerpen weergeven. Ten gevolge hiervan is weinig bekend over de reactie van het menselijke brein op individuele voorwerpen. Daarnaast is het lastig om data die zijn verkregen in mensen op een kwantitatief niveau te vergelijken met data verkregen in apen omdat deze data gebaseerd zijn op verschillende meeteenheden waarvan de correspondentie onbekend is. Soortgelijke problemen spelen een rol bij het relateren van hersendata aan computationele modellen en gedragsmetingen. Deze correspondentieproblemen bemoeilijken de ontwikkeling van een integrale theorie van visuele perceptie, en, meer in het algemeen, de ontwikkeling van een overkoepelende neurowetenschappelijke systeemtheorie. Het werk dat in dit proefschrift beschreven wordt pakt deze zaken aan door (1) het meten van de fMRI respons in IT voor plaatjes van individuele voorwerpen uit het dagelijks leven (2) het vergelijken van de zo verkregen data met data gebaseerd op onderzoek in apen, computationele modellen, en menselijk gedrag, gebruik makend van “representational similarity analysis (RSA)” dat speciaal voor dit doel ontwikkeld werd. Hieronder volgt een samenvatting van elk hoofdstuk; deze samenvattingen worden afgesloten met een compacte samenvatting van het gehele proefschrift die tevens een discussie van de belangrijkste bevindingen bevat.

Samenvattingen per hoofdstuk

In **Hoofdstuk 1** worden de verschillende fMRI analyse methoden die in dit proefschrift gebruikt worden beschreven en vergeleken. Vooral patrooninformatie analyse komt aan de orde. Dit is een analysetechniek die gedurende de laatste jaren populair is geworden in de cognitieve neurowetenschap. Doel van

deze techniek is het detecteren van verschillen in multi-voxel patronen van hersenactiviteit, behorende bij verschillende experimentele condities. Deze verschillen zijn te interpreteren als een reflectie van verschillen in de onderliggende patronen van neurale activiteit, waarvan gedacht wordt dat ze de inhoud van onze waarneming en gedachten representeren. Een andere techniek die de inhoud van representaties onderzoekt is de fMRI-adaptatietechniek. Deze techniek is gebaseerd op de redenering dat fMRI-effecten die ten gevolge van een stimulusverandering worden waargenomen in een specifiek hersengebied erop wijzen dat het hersengebied neuronen bevat die de veranderde stimuluseigenschap representeren.

De fMRI-adaptatietechniek wordt veel gebruikt in onderzoek naar gezichts-waarneming, met name om de aanwezigheid van representaties van individuele gezichten aan te tonen. Verschillende fMRI adaptatiestudies hebben de “fusiform face area” (FFA) aangewezen als de locatie waar individuele gezichten gerepresenteerd worden, maar deze studies hebben niet grondig onderzocht (1) of de gevonden effecten alleen in FFA plaats vinden en (2) wat het effect is van lagere-orde stimulusveranderingen op de spatiële omvang van de waargenomen effecten. **Hoofdstuk 2** rapporteert dat het effect van een verandering van gezichtsidentiteit niet beperkt is tot FFA: effecten werden ook gevonden in vroege visuele gebieden en in de “parahippocampal place area” (PPA). Om er zeker van te zijn dat deze effecten niet konden worden toegeschreven aan lagere-orde veranderingen van de stimulus hebben we belichting en gezichtspunt gevarieerd voor zowel identiteitsveranderingen als repetities. Dit verminderde de spatiële omvang van de effecten, maar effecten waren nog steeds aanwezig buiten gebieden die selectief op gezichten reageren. Onze bevindingen zouden geïnterpreteerd kunnen worden als bewijs voor de aanwezigheid van hogere-orde representaties van gezichtsidentiteit in vroege visuele gebieden en PPA, maar dit is onwaarschijnlijk, gegeven de al bekende functionele eigenschappen van deze hersengebieden. Mogelijke alternatieve verklaringen behelzen algemene effecten van aandacht en overdracht van activiteit van andere hersengebieden, die verbonden zijn met de onderzochte gebieden. Onze resultaten suggereren dat fMRI-effecten die ten gevolge van een stimulusverandering worden waargenomen, geen sluitend bewijs vormen voor een onderliggende neurale representatie van de veranderde stimuluseigenschap.

Naar aanleiding van de bevindingen van Hoofdstuk 2 hebben we de fMRI-adaptatietechniek gelaten voor wat hij is en zijn we overgestapt op het simpelweg direct meten van fMRI activiteit gedurende het zien van individuele voorwerpen. Hiervoor hebben we gebruikt gemaakt van “ungrouped-events designs”, waarin elk voorwerp wordt behandeld als een aparte conditie. Ungrouped-events designs worden toegepast in “representational similarity analysis” (RSA), dat geïntroduceerd wordt in **Hoofdstuk 3**. RSA is een nieuw experimen-

teel en data-analytisch raamwerk, dat het mogelijk maakt om data van verschillende takken van de cognitieve neurowetenschap op een kwantitatief niveau met elkaar te vergelijken, door te abstraheren van de activatiepatronen en de data te vergelijken op het niveau van gelijkenissen tussen activatiepatronen. Door gerepresenteerde informatie te vergelijken op het niveau van gelijkenissen tussen activatiepatronen is het niet meer nodig om de correspondentie tussen meeteenheden te bepalen. Mogelijke toepassingen van RSA zijn validatie van computationele modellen aan de hand van hersenactiviteit, vergelijking van hersenrepresentaties tussen verschillende diersoorten, analyse van connectiviteit tussen hersengebieden op het niveau van gerepresenteerde informatie, en verkenning van de representatieve inhoud van hersengebieden met behulp van ungrouped-events designs die een groot aantal experimentele condities bevatten. Statistische inferentie binnen het RSA raamwerk berust op permutatie- en bootstraptechnieken. RSA wordt gedemonstreerd door representaties van alledaagse voorwerpen in vroege visuele gebieden en FFA, gemeten met behulp van fMRI, te vergelijken met representaties van dezelfde voorwerpen in een reeks van computationele modellen. De representatie van voorwerpen in vroege visuele gebieden kon het best verklaard worden aan de hand van een eenvoudig silhouet model. De representatie in FFA kon het best verklaard worden aan de hand van een conceptueel gezicht-dier-prototype model. Deze bevindingen zijn in overeenstemming met de functionele eigenschappen van deze gebieden als gevonden in eerdere studies. Voordat we ingaan op toepassingen van RSA in de Hoofdstukken 5 en 6, worden eerst de bevindingen van Hoofdstuk 4 besproken.

Hoofdstuk 4 slaat een brug tussen conventionele fMRI-analyse en RSA. Conventionele studies hebben representaties van voorwerpen in IT onderzocht door te kijken naar de activiteit van hersengebieden als geheel, waarbij de data de activiteit gemiddeld over voorwerpen behorende bij dezelfde categorie weergaven. RSA daarentegen onderzoekt patrooninformatie voor individuele voorwerpen, waardoor twee sprongen tegelijkertijd worden gemaakt: van activiteit gemiddeld over individuele voorwerpen naar activiteit voor elk voorwerp apart en van de analyse van activiteit van een hersengebied als geheel naar de analyse van informatie gerepresenteerd in patronen van activiteit binnen een hersengebied. Dit resulteert in twee gaten die gedicht moeten worden. Het eerste gat, patrooninformatie analyse van activiteit gemiddeld over voorwerpen, is al gedicht door eerder onderzoek. Wat ontbreekt is de analyse van activiteit van hersengebieden als geheel (met name hersengebieden die selectief reageren op een bepaalde categorie van voorwerpen) voor individuele voorwerpen. Met andere woorden, houdt eerder gemeten selectiviteit voor bepaalde categorieën stand voor individuele voorwerpen? Het dichten van dit tweede gat introduceert tevens een link met neurofysiologische studies in apen, waarin het gebruikelijk is om neurale activiteit voor individuele voorwerpen te meten. We hebben

fMRI activiteit in categorie-selectieve gebieden FFA en PPA gemeten voor 96 plaatjes van alledaagse voorwerpen behorende tot een groot aantal categorieën, inclusief gezichten en huizen. We vonden geen bewijs voor de hypothese dat er plaatjes van buiten de geprefereerde categorie zouden zijn die een sterkere respons zouden veroorzaken dan plaatjes van binnen de geprefereerde categorie. Dit gold zowel voor FFA als PPA. Activiteit van een gebied als geheel lijkt dus een perfecte reflectie te geven van de categorie waartoe een waargenomen individueel voorwerp behoort. Individuele plaatjes binnen elke categorie veroorzaakten verschillende niveaus van activiteit. Dit suggereert gradatie van activiteit in plaats van een puur stapsgewijs responsprofiel. De fMRI data verkregen voor de 96 plaatjes werden vervolgens in de laatste twee hoofdstukken geanalyseerd voor patrooninformatie binnen het RSA raamwerk.

Hoofdstuk 5 vergelijkt de IT-representaties van dezelfde plaatjes van voorwerpen tussen mensen (fMRI) en apen (electrofysiologie). Onze resultaten laten zien dat activatiepatronen in IT clusters vormen, die overeenkomen met natuurlijke categorieën: levende en niet-levende voorwerpen vormen topclusters, gezichten en lichaamsdelen vormen subclusters binnen de levende voorwerpen. Deze hiërarchische categorische structuur inherent aan IT kwam overeen tussen mens en aap. Ook binnen categorieën werd een significante gelijkenis gevonden tussen de twee soorten. Deze resultaten bleven onveranderd na verwijdering van categorie-selectieve gebieden FFA en PPA van de analyse. Een uitzondering op de gelijkenis tussen mens en aap werd gevormd door de representatie van individuele gezichten: de representaties van gezichten van de eigen soort waren beter van elkaar te onderscheiden dan die van de niet-eigen soort. De categorische structuur van de representatie in IT kon niet verklaard worden aan de hand van een reeks eenvoudige en meer complexe computationele modellen. Dit suggereert dat lagere-orde visuele gelijkenis tussen plaatjes niet afdoende is om onze bevindingen te verklaren. Alles bij elkaar suggereren onze resultaten dat de codering van voorwerpen in IT tot stand komt door een categorische en een continue representatie te combineren, en dat deze gecombineerde representatie gedeeld wordt door mens en aap.

De aanwezigheid van clusters van activatiepatronen die overeenkomen met natuurlijke categorieën suggereert een link tussen de IT-representatie en menselijke waarneming. Deze link werd onderzocht in **Hoofdstuk 6** door de IT-representatie uit Hoofdstuk 5 te relateren aan paarsgewijze gelijkheidsbeoordelingen voor dezelfde 96 plaatjes. Aangezien onze stimulusset relatief groot was, zou het erg tijdrovend zijn geweest om conventionele methoden voor het verkrijgen van paarsgewijze gelijkheidsbeoordelingen te gebruiken. Daarom hebben we een nieuwe “multi-arrangement” (MA) methode ontwikkeld voor het efficiënt meten van subject-specifieke waargenomen gelijkenissen tussen grote aantallen stimuli. De zo verkregen gelijkheidsbeoordelingen van de 96 plaatjes

lieten een clustering van voorwerpen op basis van categorieën zien en deelden meerdere kenmerken met de IT-representatie, inclusief het primaire onderscheid tussen levende en niet-levende voorwerpen. Ook binnen categorieën werd een significante gelijkenis gevonden tussen IT en gelijkheidsbeoordelingen. Deze bevindingen suggereren dat hogere-orde bewuste gelijkheidsbeoordelingen van alledaagse voorwerpen gebaseerd zijn op visuele gelijkenissen en evolutionair gezien belangrijke categorieën die gerepresenteerd zijn op het niveau van IT in mens en aap. Gelijkheidsbeoordelingen stijgen echter boven de IT-representatie uit in termen van een sterkere categorische component en de introductie van soort-specifieke (mens/dier) en evolutionair recente (natuurlijk/artificieel) categorieën.

Algemene samenvatting en discussie

Onze bevindingen suggereren dat de categorie waartoe een individueel voorwerp behoort een belangrijk organisationeel principe vormt van de IT-representatie in primaten. Dit is geen nieuw idee: eerdere studies hebben aangetoond dat hersenresponsen in IT informatie bevatten over de categorie waartoe waargenomen voorwerpen behoren (o.a. Tanaka, 1996; Tsao et al., 2006; Puce et al., 1995; Kanwisher et al., 1997; Haxby et al., 2001). Wij gaan echter verder dan deze studies door het gebruik van ungrouped-events designs en RSA, en laten zien dat de structuur van de representatie van voorwerpen in IT inherent categorisch en hiërarchisch georganiseerd is, met clusters van voorwerpen die evolutionaire relevantie reflecteren, en die overeenkomen tussen mens en aap. Een reflectie van deze “primate-default” categorieën, die gerepresenteerd zijn op het niveau van IT, werd gevonden in gelijkheidsbeoordelingen van alledaagse voorwerpen. De categorische structuur van de representatie in IT kon niet verklaard worden aan de hand van een reeks eenvoudige en meer complexe computationele modellen. Onze resultaten suggereren dat modellen van IT verbeterd kunnen worden door het implementeren van sensitiviteit voor objecteigenschappen die informatief zijn voor het onderscheiden van categorieën, door middel van training onder supervisie (Ullman et al., 2002; Sigala and Logothetis, 2002).

Onze resultaten tonen aan dat IT niet alleen categorieën onderscheidt, maar ook individuele voorwerpen. Deze bevinding komt overeen met eerdere bevindingen in apen (o.a. Young and Yamane, 1992; Hung et al., 2005) en met recent gerapporteerde verschillen in patrooninformatie in IT voor individuele voorwerpen behorende tot dezelfde categorie (Kriegeskorte et al., 2007; Eger et al., 2008). Onze data laten zien dat de representationele structuur binnen categorieën een significante gelijkenis vertoont tussen mens en aap, en ook tussen hersendata en gedragsmetingen van menselijke waarneming. De gelijkenis bin-

nen categorieën wordt waarschijnlijk veroorzaakt door visuele overeenkomsten en verschillen tussen voorwerpen die tot dezelfde categorie behoren, en is consistent met de eerder gerapporteerde relatie tussen waargenomen vorm van voorwerpen en activatiepatroon gelijkenissen in IT (Edelman et al., 1998; Haushofer et al., 2008; Op de Beeck et al., 2001; 2008).

De sterkste clustering van activatiepatronen in zowel IT als FFA werd gevonden voor menselijke gezichten, waarschijnlijk ten gevolge van de hoge mate van visuele gelijkenis tussen verschillende individuele gezichten. Overeenkomend met deze observatie rapporteerde een eerdere studie dat geen detecteerbare patrooninformatie over gezichtsidentiteit aanwezig was in FFA. Deze informatie werd wel gevonden in de anterieure IT cortex (Kriegeskorte et al., 2007). Electrofysiologische metingen van de “middle macaque face patch”, een hersengebied in apen dat functioneel homoloog lijkt te zijn aan FFA, toonden aan dat de activiteit van populaties neuronen in dit gebied informatie over gezichtsidentiteit bevat (Tsao et al., 2006), maar het grootste deel van de variatie in respons werd verklaard door effecten van categorie. Een recente fMRI studie in mensen toonde de aanwezigheid van informatie over gezichtsidentiteit aan in de anterieure FFA en IT met behulp van spatiële-temporele patrooninformatie analyse (Nestor et al., 2011). Onze bevindingen van Hoofdstuk 2 lijken overeen te komen met deze recente bevindingen, maar kunnen geen sluitend bewijs leveren voor de aanwezigheid van representaties van individuele gezichten in FFA. Samenvattend: het huidige bewijs suggereert dat gezichtsherkenning wordt uitgevoerd door een netwerk van hersengebieden, waaronder FFA en de anterieure IT. Responsen in FFA bevatten misschien een bepaalde hoeveelheid informatie over de identiteit van waargenomen gezichten, maar de belangrijkste functie van dit gebied lijkt detectie van gezichten te zijn (Puce et al., 1995; Kanwisher et al., 1997; Kriegeskorte et al., 2007) en niet identificatie.

Onze bevindingen suggereren dat IT in primaten een code bevat die gedeeld wordt door mens en aap, en dat deze code tot stand komt door de combinatie van een categorische en een continue representatie van voorwerpen. Deze code zou geïmplementeerd kunnen zijn als een continue “kaart” van voorwerpeigenschappen met meerdere clusters van gerelateerde eigenschappen, die afgestemd zijn op het discrimineren van categorieën met een hoge mate van gedragsmatige en evolutionaire relevantie (bv. gezichten) (Haxby et al., 2001). Deze kaart zou het bestaan van categorie-selectieve gebieden kunnen verklaren, en zou consistent zijn met onze bevinding dat activiteit van deze gebieden een perfecte reflectie lijkt te geven van de categorie waartoe een waargenomen individueel voorwerp behoort, maar tegelijkertijd ook gradatie laat zien in plaats van een puur stapsgewijs responsprofiel. Daarnaast zou zo’n kaart consistent zijn met onze bevinding dat activatiepatronen clusters vormen die overeenkomen met natuurlijke categorieën, ondanks het ontbreken van stapsgewijze categorische respon-

sen in IT op het niveau van individuele neuronen of voxels. Stapsgewijze categorische responsen zijn wel gerapporteerd voor neuronen in de prefrontale cortex (Ashby and Ell, 2001; Freedman et al., 2001). De prefrontale cortex ontvangt informatie van IT en combineert mogelijk de patrooninformatie uit IT om scheidingen tussen categorieën te verduidelijken op een taakafhankelijke, flexibele wijze. De prefrontale cortex draagt misschien, in combinatie met IT, ook bij aan hogere-orde bewuste gelijkheidsbeoordelingen van voorwerpen. Richtingen voor toekomstig onderzoek behelzen nader onderzoek naar de aard van de kaart(en) in IT met behulp van hoogveld fMRI (zie Op de Beeck et al., 2008) en het exploreren van de representatieve connectiviteit tussen IT en de prefrontale cortex met behulp van RSA.